ASYMPTOTIC PROPERTIES OF THE NONPARAMETRIC PART IN PARTIAL LINEAR HETEROSCEDASTIC REGRESSION MODELS

Hua Liang, Wolfgang Härdle and Axel Werwatz *

Abstract

This paper considers estimation of the unknown function $g(\bullet)$ in the partial linear regression model $Y_i = X_i^T\beta + g(T_i) + \varepsilon_i$ with heteroscedastic errors. We first construct a class of estimates g_n of g and prove that, under appropriate conditions, g_n is weak, mean square error consistent. Rates of convergence and asymptotic normality for the estimator g_n are also established.

Key Words and Phrases:Key words and phrases: Asymptotic normality, consistency, heteroscedasticity, kernel estimation, rates of convergence, partial linear model, semiparametric models.

1 INTRODUCTION

Semiparametric models combine the flexibility of nonparametric modeling with structural parametric components. One such model that has received a lot of attention in the literature is the semiparametric partial linear regression model

$$Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, i = 1, \ldots, n \tag{1}$$

*Hua Liang is Associate Professor of Statistics, Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China. Wolfgang Härdle is Professor of Econometrics, Axel Werwatz is Dr. of Economics. Both of latter are at the Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany. This research was supported by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" . The first author was supported by Alexander von Humboldt Foundation. The authors would like to thank Mr. Knut Bartels for his valuable comments which greatly improved the presentation of this paper.

where X and T are (possibly) multidimensional regressors, β a vector of unknown parameters, $g(\bullet)$ an unknown smooth function and ε an error term with mean zero conditional on X and T.

Well-known applications in the econometrics literature that can be put in the form of (1) are the human capital earnings function (Willis (1986)) and the wage curve (Blanchflower and Oswald (1994)). In both cases, log-earnings of an individual are related to personal characteristics (sex, marital status) and measures of a person's human capital like schooling and labor market experience. Economic theory suggests a non-linear relationship between log-earnings and labor market experience, which therefore plays the role of the variable T in (1). The wage curve is obtained by including the local unemployment rate as an additional regressor, with a possibly non-linear influence. Rendtel and Schwarze (1995), for instance, estimate $g(\bullet)$ as a function of the local unemployment rate using smoothing-splines and find a U-shaped relationship.

Under various assumptions, several authors have considered estimation of β in (1) at a parametric rate. Chen (1988), Heckman (1986), Robinson (1988), Schick (1996) and Speckman (1988) constructed $\sqrt{n}-$consistent estimators of β. Cuzick (1992a) studied efficient estimation of β when the error density is known. Efficient estimation when the error distribution is of an unknown form is treated in Cuzick (1992b) and Schick (1993).

In this paper, we will instead focus on deriving the asymptotic properties of an estimator of the unknown function $g(\bullet)$. We consider its consistency, weak convergence rate and asymptotic normality. We will derive these results for a specific version of (1) with nonstochastic regressors, heteroscedastic errors and T univariate.

The remainder of this paper is organized as follows. In the following section we will describe methods for estimating β and $g(\bullet)$. We prove consistency and asymptotic normality of the estimator of $g(\bullet)$ in sections 3 and 4. We illustrate the usefulness of the estimator and the relevance of the asymptotic distribution results for applied work by a small-scale Monte Carlo study and an empirical illustration in the final section of the paper.

2 THE ESTIMATOR

Specifically, we consider estimation of $g(\bullet)$ (and β) in the following partial linear, semiparametric regression model:

$$Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, i = 1, \ldots, n \tag{2}$$

where β is an unknown p dimensional parameter vector, $g(\bullet)$ an unknown, smooth function from $[0,1]$ to $\mathbb{R}^1$, (X_1, T_1), (X_2, T_2) ... are known, nonrandom design points and $\varepsilon_1, \ldots, \varepsilon_n$ are independent mean zero random errors with nonconstant variance. We allow the variance of ε to depend on X and T in an arbitrary way.

Previous work in a heteroscedastic setting has focused on the nonparametric regression model (i.e. $\beta = 0$). Müller et al. (1987) proposed an estimate of the variance function by using kernel smoother, and then proved that the estimate is uniformly consistent. Hall and Carroll (1989) considered consistency of estimates of $g(\bullet)$. Eubank et al. (1990) proposed trigonometric series type estimators g_λ of g. They investigated asymptotic approximations of the integrated mean squared error and the partial integrated mean squared error of g_λ. The heteroscedastic version of (1) with $\beta \neq 0$ has been considered in Schick (1996) but he considers weighted least squares estimation of β. We focus on nonparametric estimation of $g(\bullet)$ as a function of T.

Suppose we knew β. Then we may estimate $g(\bullet)$ by nonparametric regression of $Y_i - X_i^T\beta$ (the variation in Y_i not accounted for by the linear component $X_i^T\beta$) on T_i.

In the literature one can find various methods for estimating $g(\bullet)$ nonparametrically, e.g., kernel, nearest neighbor, orthogonal series, piecewise polynomial and smoothing splines. See Härdle (1990) for an extensive discussion of their statistical properties. All these estimators may be written as weighted local averages of the observed values of the dependent variable with the weights depending on the values of the explanatory variables. In our case, we can write (still assuming that β is known):

$$\hat{g}(t) = \sum_{i=1}^{n} \omega_{ni}(t)(Y_i - X_i^T\beta), \tag{3}$$

where $\omega_{ni}(t) = \omega_{ni}(t; T_1, T_2, \ldots, T_n)$ are weight functions that depend on the observations $T_1, \ldots, T_n$.

For instance, a Gasser-Müller-type kernel estimator takes

$$\omega_{ni}(t) = \frac{1}{h}\int_{s_{i-1}}^{s_i} K\Big(\frac{t-s}{h_n}\Big)ds \quad 1 \leq i \leq n$$

for $s_0 = 0$, $s_n = 1$, $s_i = \frac{1}{2}(T_{(i)} + T_{(i+1)})$. Here $T_{(1)}, \ldots, T_{(n)}$ are sample order statistics, $K(\bullet)$ is the kernel function and h denotes the bandwidth. See Remark 5 below for details.

Given the estimator $\hat{g}(t)$ as defined in (3) we may estimate β by the least squares regression of

$$Y_i \;=\; X_i^T\beta + \hat{g}(T_i) + \epsilon_i$$

$$\begin{aligned} Y_i - \sum_{j=1}^{n} \omega_{nj}(T_i) Y_j &= \Big\{ X_i - \sum_{j=1}^{n} \omega_{nj}(T_i) X_j \Big\}^T \beta + \epsilon_i \\ \widetilde{Y}_i &= \widetilde{X}_i^T \beta + \epsilon_i \end{aligned} \tag{4}$$

That is, we estimate β by the generalized least squares estimator

$$\beta_{LS} = (\widetilde{X}^T \widetilde{X})^{-1} \widetilde{X}^T \widetilde{Y} \tag{5}$$

where $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_n)^T$ and $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)^T$ are the presmoothed design and response variables.

In the final step we obtain the feasible estimator of $g(\bullet)$ by substituting β_{LS} for the unknown β in (3):

$$g_n(t) = \sum_{i=1}^{n} \omega_{ni}(t)(Y_i - X_i^T \beta_{LS}), \tag{6}$$

Further motivation for the estimators defined in (5) and (6) is given in Speckman (1988) and Gao, et al. (1995). Note though that β_{LS} is not an efficient estimator in the sense of asymptotic normality.

In the following section we state and prove the weak, mean square error consistency and give the rates of convergence of $g_n(t)$ under various assumptions.

3 CONSISTENCY RESULTS

All technical preliminaries needed in the proofs of the following results are collected in Appendix 6 as lemmas. For convenience and simplicity, we always let C denote some positive constant not depending on n. We will use the following assumptions.

Assumption 1. *There exist continuous functions $h_j(\bullet)$ defined on $[0,1]$ such that each element of X_i satisfies*

$$x_{ij} = h_j(T_i) + u_{ij} \quad 1 \le i \le n, \quad 1 \le j \le p \tag{7}$$

where u_{ij} is a sequence of real numbers which satisfy $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_i = 0$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_i u_i^T = B \tag{8}$$

is a positive definite matrix, and

$$\limsup_{n\to\infty} \frac{1}{a_n} \max_{1\le k\le n} \left| \sum_{i=1}^{k} u_{jim} \right| < \infty \quad \textit{for } m = 1, \ldots, p \tag{9}$$

holds for all permutations $(j_1, \ldots, j_n)$ *of* $(1, 2, \ldots, n)$ *where* $u_i = (u_{i1}, \ldots, u_{ip})^T$, $a_n = n^{1/2} \log n$.

Assumption 2.

(a) $\sum_{i=1}^n \omega_{ni}(t) \to 1$ as $n \to \infty$;

(b) $\sum_{i=1}^n |\omega_{ni}(t)| \le C$ for all t and some constant C;

(c) $\sum_{i=1}^n |\omega_{ni}(t)| I(|t - T_i| > a) \to 0$ as $n \to \infty$ for all $a > 0$;

(d) $\sup_{i \le n} |\omega_{ni}(t)| = O\{(\log n)^{-1}\}$.

Denote $n_t = \left\{\sum_{i=1}^n \omega_{ni}^2(t)\right\}^{-1}$.

Assumption 3.

(a) $\sup_n (n_t) \sup_{1 \le i \le n} |\omega_{ni}(t)| < \infty$, and $n_t = o(n)$;

(b) $\sqrt{n_t} \sup_{1 \le i \le n} |\omega_{ni}(t)| = O(n^{-\alpha/2})$ for some $1 > \alpha > 0$;

(c) $\sum_{i=1}^n \omega_{ni}^2(t) E\varepsilon_i^2 = \sigma_0^2/n_t + o(1/n_t)$ for some $\sigma_0^2 > 0$.

Remark 1. *Assumption 1 is a common requirement for proving consistency of* β *in the partial linear model (1). In fact, (7) of Assumption 1 is parallel to the case*

$$h_j(T_i) = E(x_{ij}|T_i) \qquad \textit{and } u_{ij} = x_{ij} - E(x_{ij}|T_i)$$

when (X_i, T_i) *are random variables. (8) is similar to the result of the strong law of large numbers for random errors. (9) is similar to law of the iterated logarithm. More detailed discussions may be found in Speckman (1988) and Gao et al. (1995).*

Theorem 1. *Under Assumptions 1 and 2*

$E\{g_n(t)\} \to g(t)$ *as* $n \to \infty$ *at every continuity point of the function* g.

Proof. Decompose the difference $g_n(t) - g(t)$ as follows by direct calculation.

$$g_n(t) - g(t) = \sum_{j=1}^n \omega_{nj}(t)\{g(T_i) + \varepsilon_i - X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T) - X_j^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon}\} - g(t)$$

where $\widetilde{g}(T) = \{\widetilde{g}(T_1), \ldots, \widetilde{g}(T_n)\}^T$ and $\widetilde{g}(T_i) = g(T_i) - \sum_{j=1}^n \omega_{nj}(T_i)g(T_j)$ and $\widetilde{\varepsilon}$ just like $\widetilde{X}$. It follows that

$$E\{g_n(t)\} - g(t) = \Big\{\sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t)\Big\} - \sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T) \qquad (10)$$

The first term tends to zero by Lemma A.1 (i). By lemmas A.2 and A.1 (i) and Cauchy-Schwarz inequality, we know that every element of $(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T)$ is $o(n^{-1/2})$, i.e.,

$$\Big\{(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T)\Big\}_j = o(n^{-1/2}) \quad \text{for } j = 1,\ldots,p \tag{11}$$

It suffices to show that every element of $\sum_{i=1}^n \omega_{ni}(t)X_i$ is $O(n^{1/2})$. Observe that

$$\sum_{i=1}^n \omega_{ni}(t)x_{ij} = \sum_{i=1}^n \omega_{ni}(t)\{h_j(T_i) + u_{ij}\}$$

Since $h_j(\bullet)$ is continuous, $\sum_{i=1}^n \omega_{ni}(t)h_j(T_i)$ converges to $h(t)$ on the continuity point of $h(t)$ by the same proof as one for Lemma A.1 (i). Moreover, by Abel's inequality and Assumption 2 (d),

$$\Big|\sum_{i=1}^n \omega_{ni}(t)u_{ij}\Big| \le \sup_{1\le i\le n} |\omega_{ni}(t)| \max_{1\le k\le n}\Big|\sum_{i=1}^k u_{j_i m}\Big| = O(n^{1/2})$$

Thus

$$\sum_{i=1}^n \omega_{ni}(t)x_{ij} = O(n^{1/2}) \tag{12}$$

and we complete the proof of Theorem 1. #

Theorem 1 shows that $g_n(t)$ is an asymptotically unbiased estimator of $g(t)$ at every continuity point of $g(t)$. The next result, Theorem 2, will demonstrate that $g_n(t)$ is also mean square-error consistent.

Theorem 2 *Assume the conditions of Theorem 1 hold except Assumption 2 (d) which is replaced by* $\sup_{i\le n}|\omega_{ni}(t)| = o\{(\log n)^{-1}\}$. *Then* $E\{g_n(t)-g(t)\}^2 \to 0$ *as* $n\to\infty$.

Proof. It follows from the C_r−inequality with $r = 2$ that

$$\begin{aligned} E\{g_n(t)-g(t)\}^2 &\le CE\Big|\sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t)\Big|^2 + CE\Big|\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i\Big|^2 \\ &+ CE\Big|\sum_{i=1}^n \omega_{ni}(t)(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T)\Big|^2 + CE\Big|\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\varepsilon\Big|^2 \end{aligned} \tag{13}$$

In the proof of Theorem 1 we obtained that the first and third terms of (13) converge to zero as n tends to infinity. The second can be shown to be order $o(1)$ by direct calculation.

We shall now prove the fourth term also converges to zero. Denote $(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T = (\eta_{ji})_{p\times n}$.

$$\begin{aligned} E\Big|\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\varepsilon\Big|^2 &= E\Big\{\sum_{i=1}^n \omega_{ni}(t)\Big(\sum_{k=1}^p\sum_{l=1}^n x_{ik}\eta_{kl}\varepsilon_l\Big)\Big\}^2 \\ &= \sum_{l=1}^n\Big(\sum_{i=1}^n\sum_{k=1}^p \omega_{ni}(t)x_{ik}\eta_{kl}\Big)^2\sigma_l^2 \end{aligned}$$

It follows from the arguments for (12) that this equals to $O(n)\sum_{l=1}^n \eta_{kl}^2$. Since $\sum_{l=1}^n \eta_{kl}^2$ and the elements of the $k-$th row of $(\widetilde{X}^T\widetilde{X})^{-1}$ have the same order $O(n^{-1})$. It follows that

$$E\Big|\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\varepsilon\Big|^2 = o(1). \tag{14}$$

Furthermore, we can easily show that

$$E\Big|\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\sum_{k=1}^n \widetilde{X}_k\Big\{\sum_{l=1}^n \omega_{nl}(T_k)\varepsilon_k\Big\}\Big|^2 = o(1). \tag{15}$$

Combining (14) and (15) ensures that the fourth term of (13) is $o(1)$, and thus completes the proof of Theorem 2. #

The following result gives the weak convergence rate of g_n under stronger assumptions on $\{\omega_{ni}(t)\}$ than those given in Assumption 2. Here we list these assumptions.

Assumption 2'. The weight functions $\omega_{ni}(t)$ satisfy:

(a) $\sup_t |\sum_{i=1}^n \omega_{ni}(t) - 1| = O_P(n^{-1/3}\log n)$;

(b) $\sup_t \sum_{i=1}^n |\omega_{ni}(t)| I(|t-T_i| > c_n) = O(d_n)$, where d_n and c_n are $n^{-1/3}\log n$;

(c) $\sup_t \max_{1\le i\le n} |\omega_{ni}(t)| = O_P(n^{-2/3})$.

Theorem 3. *Assume $g(\bullet)$ and $h_j(\bullet)$ are Lipschitz continuous of order 1 and Assumptions 1 and 2' hold. Then*

$$g_n(t) - g(t) = O_P(n^{-1/3}\log n).$$

Proof. By Lemma A.1 (ii),

$$\sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t) = O_P(n^{-1/3}\log n).$$

Using Assumption 2' (c) and Chebyshev's inequality we have

$$\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i = O_P(n^{-1/3}\log n).$$

The similar arguments as that for (11) and (12) yield

$$\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{g}(T) = O_P(n^{-1/3}\log n).$$

Finally, observe that $\sum_{i=1}^n \omega_{ni}(t)u_{ij} = O(1)$ and then $\sum_{i=1}^n \omega_{ni}(t)x_{ij} = O(1)$ for $j = 1, \ldots, p$. Thus, by the arguments for (14) and (15),

$$E\Big|\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon}\Big|^2 = O(n^{-1}). \tag{16}$$

which entails

$$\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon} = O_P(n^{-1/3}\log n).$$

This completes the proof of Theorem 3. #

Remark 2. *We can conclude from the above arguments that*

$$\limsup_{n\to\infty}(n^{2/3}\log^{-2} n)E\{g_n(t) - g(t)\}^2 < \infty.$$

Theorem 4 gives the asymptotic variance of $g_n(t)$.

Theorem 4. *Under Assumptions 1, 2' and 3, $n_t Var\{g_n(t)\} \to \sigma_0^2$ as $n \to \infty$.*

Proof.

$$\begin{aligned} n_t Var\{g_n(t)\} &= n_t E\Big\{\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i\Big\}^2 + n_t E\Big\{\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon}\Big\}^2 \\ &\quad -2n_t E\Big\{\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i\Big\}\cdot\Big\{\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon}\Big\} \end{aligned}$$

The first term converges to σ_0^2. The second term tends to zero by (16), and then the third term also tends to zero by the Cauchy-Schwarz inequality. #

4 ASYMPTOTIC NORMALITY

In the nonparametric regression model, Liang (1995) proved asymptotic normality for independent ε_i's under the mild conditions. In this section, we shall consider the asymptotic normality of g_n under the Assumptions 1, 2' and 3.

Theorem 5. *Assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent random variables with $E\varepsilon_i = 0$ and $\inf_i \sigma_i^2 > c_\sigma > 0$ for some c_σ. There exists a function $G(u)$ satisfying*

$$\int_0^\infty uG(u)du < \infty \tag{17}$$

such that

$$P(|\varepsilon_i| > u) \le G(u), \quad \text{for } i = 1, \ldots, n \quad \text{ and large enough } u. \tag{18}$$

If

$$\frac{\max_{1\le i\le n}\omega_{ni}^2(t)}{\sum_{i=1}^n \omega_{ni}^2(t)} \to 0 \qquad \text{as} \quad n\to\infty. \tag{19}$$

Then

$$\frac{g_n(t)-Eg_n(t)}{\sqrt{Var\{g_n(t)\}}} \longrightarrow^{\mathcal{L}} N(0,1) \quad \text{as } n\to\infty.$$

Remark 3. *The condition* $\int_0^\infty uG(u)du<\infty$ *is to guarantee* $\sup_i \sigma_i^2<\infty$.

The proof of Theorem 5. At first from the proof of Theorem 4, we obtain that

$$Var\{g_n(t)\} = \sum_{i=1}^n \omega_{ni}^2(t)\sigma_i^2 + o\Big\{\sum_{i=1}^n \omega_{ni}^2(t)\sigma_i^2\Big\}$$

Furthermore

$$g_n(t)-Eg_n(t)-\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i = \sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon} = O_P(n^{-1/2})$$

which yields

$$\frac{\sum_{i=1}^n \omega_{ni}(t)X_i^T(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{\varepsilon}}{\sqrt{Var\{g_n(t)\}}} = O_P(n^{-1/2}n_t^{1/2}) = o_P(1)$$

It follows that

$$\frac{g_n(t)-Eg_n(t)}{\sqrt{Var\{g_n(t)\}}} = \frac{\sum_{i=1}^n \omega_{ni}(t)\varepsilon_i}{\sqrt{\sum_{i=1}^n \omega_{ni}^2(t)\sigma_i^2}} + o_P(1) \stackrel{\text{def}}{=} \sum_{i=1}^n a_{ni}^*\varepsilon_i + o_P(1),$$

where $a_{ni}^* = \frac{\omega_{ni}(t)}{\sqrt{\sum_{i=1}^n \omega_{ni}^2(t)\sigma_i^*}}$. Let $a_{ni}=a_{ni}^*\sigma_i$, obviously $\sup_i\sigma_i<\infty$ due to $\int_0^\infty vH(v)dv<\infty$. The proof of the theorem immediately follows from the conditions (17)–(19) and Lemma A.4. #

Remark 4. *If* $\varepsilon_1,\ldots,\varepsilon_n$ *are independent identically distributed, then* $E|\varepsilon_1|^2<\infty$ *and the condition (19) of Theorem 5 can yield the result of Theorem 5.*

Assumption 2". *The weight functions* $\omega_{ni}(t)$ *satisfy:*

(a) $\sum_{i=1}^n \omega_{ni}(t)-1=o(n_t^{-1/2})$;

(b) $\sum_{i=1}^n |\omega_{ni}(t)|I(|t-T_i|>c_n')=o(d_n')$, *where* c_n' *and* d_n' *are* $o(n_t^{-1/2})$.

Theorem 6. *Suppose that $g(t)$ is Lipschitz continuous of order 1. Assume the conditions of Theorem 5 hold with the previous Assumption 2" replacing Assumption 2'. Then*

$$\frac{g_n(t) - g(t)}{\sqrt{Var\{g_n(t)\}}} \longrightarrow^{\mathcal{L}} N(0,1) \text{ as } n \to \infty.$$

Proof. In fact, recall the conclusion of Theorem 5, it suffices to show that

$$\frac{Eg_n(t) - g(t)}{\sqrt{Var\{g_n(t)\}}} = \sqrt{n_t}\{Eg_n(t) - g(t)\} + o(1) = o(1).$$

For $c_n' = o(n_t^{-1/2})$. Note that

$$\begin{aligned} |Eg_n(t) - g(t)| &\leq \sum_{i=1}^n |\omega_{ni}(t)\{g(T_i) - g(t)\}|\{I(|T_i - t| > c_n') + I(|T_i - t| \leq c_n')\} \\ &\quad + |g(t)|\Big|\sum_{i=1}^n \omega_{ni}(t) - 1\Big| \\ &\leq \delta(g, c_n') \cdot B + 2C\sum_{i=1}^n |\omega_{ni}(t)| I(|T_i - t| > c_n') + C\Big|\sum_{i=1}^n \omega_{ni}(t) - 1\Big|, \end{aligned}$$

where $C = \sup_{t\in[0,1]} |g(t)|$ and $\delta(g, c_n') = \sup_{|t-t'|\leq c_n'} |g(t) - g(t')|$. Assumption 2" and the previous arguments yield the conclusion of Theorem 6. #

Remark 5. *In this remark, we shall give concrete weight functions $\{\omega_{ni}(t), i = 1, \ldots, n\}$ which satisfy the assumptions given in the former context, in order to explain the reasonability of the results established in previous sections carefully.*

Assume

$$\omega_{ni}(t) = \frac{1}{h_n}\int_{s_{i-1}}^{s_i} K\Big(\frac{t-s}{h_n}\Big)ds \quad 1 \leq i \leq n \tag{20}$$

where $s_0 = 0$, $s_n = 1$ and $s_i = \frac{1}{2}(T_{(i)} + T_{(i+1)})$, $1 \leq i \leq n-1$. h_n is a sequence of bandwidth parameters which tends to zero as $n \to \infty$ and $K(\bullet)$ is a kernel function, which is supported to have compact support and to satisfy

$$\text{supp}(K) = [-1,1], \sup|K(x)| \leq C < \infty, \int K(u)du = 1 \text{ and } K(u) = K(-u)$$

Obviously Assumptions 2(a), (b) and (d) are satisfied for the weight functions given in (20). If

$$\int_{|u|\geq ah_n^{-1}} K(u)du = o(1)$$

Then Assumption 2 (c) hold also. In fact

$$\begin{aligned}\sum_{i=1}^{n}\omega_{ni}(t)I(|T_i-t|>a) &= \frac{1}{h_n}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i}K\Big(\frac{t-s}{h_n}\Big)dsI(|T_i-t|>a)\\ &\le \frac{1}{h_n}\int_{|T_i-s|\ge a-\max|T_i-T_{i-1}|}K\Big(\frac{t-s}{h_n}\Big)ds\\ &\le \frac{1}{h_n}\int_{|u|\ge h_n^{-1}(a-\max|T_i-T_{i-1}|)}K(u)du=o(1)\end{aligned}$$

Now let us take $h_n = Cn^{-1/3}$ for some $C>0$ and suppose

$$\int_{|u|\ge ah_n^{-1}}K(u)du = O(n^{-1/3}\log n)$$

There exist constants $C_1, C_2 > 0$ such that

$$\frac{C_1}{n}\le \min_{1\le i\le n}|T_i-T_{i-1}|\le \max_{1\le i\le n}|T_i-T_{i-1}|\le \frac{C_2}{n}$$

Then we can take $n_t = nh_n$, and Assumptions 3 and 2" hold. Theorems 5 and 6 imply that

$$\sqrt{nh_n}\{g_n(t)-g(t)\}\longrightarrow^{\mathcal{L}} N(0,\sigma_0^2) \text{ as } n\to\infty.$$

This is just the classical conclusion in nonparametric regression estimation.

5 NUMERICAL EXAMPLES

In this section we will illustrate the finite-sample behaviour of the estimator by applying it to real data and by performing a small simulation study.

In the introduction we already mentioned the human-capital earnings function as a well-known econometric application that can be put into the form of a partial linear model. It typically relates the logarithm of earnings to a set of explanatory variables describing an individual's skills, personal characteristics and labour market conditions. Specifically, we estimate β and $g(\bullet)$ in the model

$$\ln Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, \tag{21}$$

where X contains two dummy variables indicating the level of secondary schooling a person has completed and T, is a measure of labour market experience (defined as the number of years spent in the labour market and approximated by subtracting (years of schooling + 6) from a person's age).

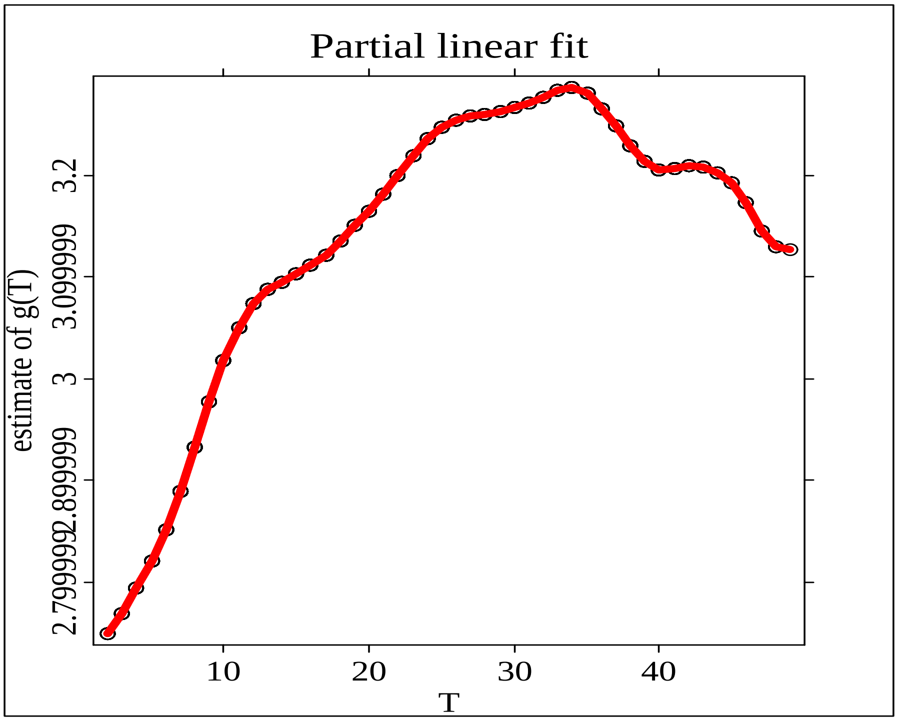

Figure 1: *Relationship of log-earnings and labour-market experience*

Under certain assumptions, the estimate of β can be interpreted as the rate of return from obtaining the respective level of secondary schooling. Regarding $g(T)$, human capital theory suggests a concave form: rapid human capital accumulation in the early stage of one's labor market carrer are associated with rising earnings that peak somewhere during midlife and decline thereafter as hours worked and the incentive to invest in human capital decrease. To allow for concavity, parametric specifications of the earnings-function typically include T and T^2 in the model and obtain a positive estimate for the coefficent of T and a negative estimate for the coefficient of T^2.

For nonparametric fitting, we use a Nadaraya-Watson weight function with quartic kernel

$$(15/16)(1-u^2)^2 I(|u| \leq 1)$$

and chose the bandwidth using cross-validation. The estimate of $g(T)$ is depicted in Figure 1. In a sample size that is lower than in most empirical investigations of the human capital earnings function we obtain an estimate that nicely agrees with the concave relationship envisioned by economic theory and often confirmed by parametric model fitting.

We also conducted a small simulation study to get fiurther insights into the small-sample

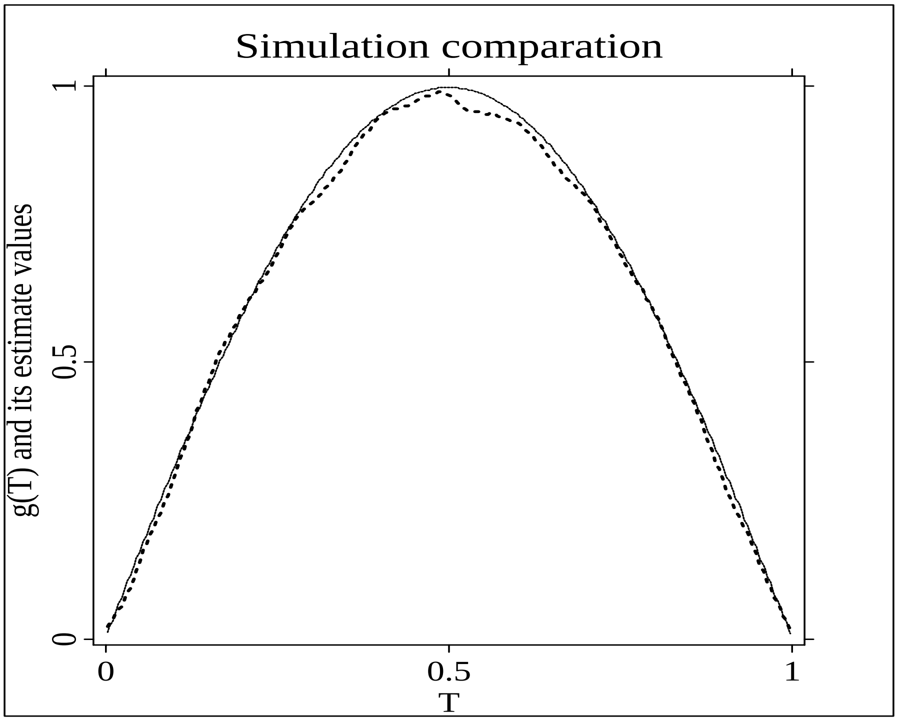

Figure 2: *Estimates of the function $g(T)$*

performance of the estimator of $g(\bullet)$. We consider the model

$$Y_i = X_i^T \beta + \sin(\pi T_i) + \sin(X_i^T \beta + T_i)\varepsilon_i, \quad i = 1, \ldots, n = 300$$

where ε_i is standard normally distributed and X_i and T_i are sampled from a uniform distribution on $[0, 1]$. We set $\beta = (1, 0.75)^T$ and performed 100 replications of generating samples of size $n = 300$ and estimating $g(\bullet)$. Figure 2 depicts the "true" curve $g(T) = \sin(\pi T)$ (solid-line) and an average of the 100 estimates of $g(\bullet)$ (dashed-line). The average estimate nicely captures the shape of $g(\bullet)$.

6 APPENDIX

In this appendix we state some useful lemmas.

Lemma A.1. *Suppose that Assumption 2 (a)-(c) hold and $g(\bullet)$ and $h_j(\bullet)$ are continuous. Then*

$$(i) \qquad \max_{1 \le i \le n} \Big| G_j(T_i) - \sum_{k=1}^{n} \omega_{nk}(T_i) G_j(T_k) \Big| = o(1)$$

Furthermore, if $g(\bullet)$ and $h_j(\bullet)$ are Lipschitz continuous of order 1 and Assumption 2' (a)-(c) and 2 (b) hold. Then

$$(ii) \qquad \max_{1\le i\le n}\Big|G_j(T_i)-\sum_{k=1}^{n}\omega_{nk}(T_i)G_j(T_k)\Big| = O(c_n+d_n)$$

for $j=0,\ldots,p$. Where $G_0(\bullet)=g(\bullet)$ and $G_l(\bullet)=h_l(\bullet)$ for $l=1,\ldots,p$.

Proof. We only present the proof of (ii) for $g(\bullet)$. The proofs of other cases and (i) are similar. Observe that

$$\begin{aligned}\sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\} &= \sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\}+\Big\{\sum_{i=1}^{n}\omega_{ni}(t)-1\Big\}g(t)\\ &= \sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\}I(|T_i-t|>c_n)\\ &\quad+\sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\}I(|T_i-t|\le c_n)+\Big\{\sum_{i=1}^{n}\omega_{ni}(t)-1\Big\}g(t)\end{aligned}$$

By Assumption 2'(b) and Lipschitz continuity of $g(\bullet)$

$$\sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\}I(|T_i-t|>c_n)=O(d_n), \tag{22}$$

and

$$\sum_{i=1}^{n}\omega_{ni}(t)\{g(T_i)-g(t)\}I(|T_i-t|\le c_n)=O(c_n). \tag{23}$$

(22)-(23) and Assumption 2 (a) complete the proof of Lemma A.1.

Lemma A.2. *Under Assumptions 1 and 2'.*

$$\lim_{n\to\infty}\frac{1}{n}\widetilde{X}^T\widetilde{X}=B$$

Proof. Denote $\bar{h}_{ns}(T_i)=h_s(T_i)-\sum_{k=1}^{n}\omega_{nk}(T_i)x_{ks}$. It follows from $x_{is}=h_s(T_i)+u_{is}$ that the (s,m) element of $\widetilde{X}^T\widetilde{X}$ $(s,m=1,\ldots,p)$ is

$$\begin{aligned}\sum_{i=1}^{n}\widetilde{x}_{is}\widetilde{x}_{im} &= \sum_{i=1}^{n}u_{is}u_{im}+\sum_{i=1}^{n}\bar{h}_{ns}(T_i)u_{im}\\ &\quad+\sum_{i=1}^{n}\bar{h}_{nm}(T_i)u_{is}+\sum_{i=1}^{n}\bar{h}_{ns}(T_i)\bar{h}_{nm}(T_i)\\ &\stackrel{\text{def}}{=} \sum_{i=1}^{n}u_{is}u_{im}+\sum_{q=1}^{3}R_{nsm}^{(q)}\end{aligned}$$

The strong law of large number implies that $\lim_{n\to\infty}1/n\sum_{i=1}^{n}u_iu_i^T=B$ and Lemma A.1 means $R_{nsm}^{(3)}=o(n)$, which and Cauchy-Schwarz inequality show that $R_{nsm}^{(1)}=o(n)$ and $R_{nsm}^{(2)}=o(n)$. This completes the proof of the lemma.

The following Lemma is a slight version of Theorem 9.1.1 of Chow and Teicher (1988). We therefore do not give a proof.

Lemma A.3. *Let* $\xi_{nk}, k = 1, \ldots, k_n$, *be independent random variables with* $E\xi_{nk} = 0$, *and* $E\xi_{nk}^2 = \sigma_{nk}^2 < \infty$. *Assume that* $\lim_{n\to\infty} \sum_{k=1}^{k_n} \sigma_{nk}^2 = 1$ *and* $\max_{1\le k\le k_n} \sigma_{nk}^2 \to 0$. *Then* $\sum_{k=1}^{k_n} \xi_{nk} \to^{\mathcal{L}} N(0,1)$ *in distribution if and only if*

$$\sum_{k=1}^{k_n} E\xi_{nk}^2 I(|\xi_{nk}| > \delta) \to 0 \quad \textit{for any } \delta > 0 \ \textit{ as } n \to \infty.$$

Lemma A.4. *Let* $V_1, \ldots, V_n$ *be independent random variables with* $EV_i = 0$ *and* $\inf_i EV_i^2 > C > 0$ *for some constant number* C. *The function* $H(v)$ *satisfying* $\int_0^\infty vH(v)dv < \infty$ *such that*

$$P\{|V_k| > v\} \le H(v) \quad \textit{for large enough } v > 0 \quad \textit{and } k = 1, \ldots, n. \tag{24}$$

Also assume that $\{a_{ni}, i = 1, \ldots, n, n \ge 1\}$ *is a sequence real numbers satisfying* $\sum_{i=1}^n a_{ni}^2 = 1$. *If*
$\max_{1\le i\le n} |a_{ni}| \to 0$, *then for* $a'_{ni} = a_{ni}/\sigma_i(V)$,

$$\sum_{i=1}^n a'_{ni} V_i \longrightarrow^{\mathcal{L}} N(0,1) \quad \text{as } n \to \infty.$$

Proof. Denote $\xi_{nk} = a'_{nk} V_k, k = 1, \ldots, n$. We have $\sum_{k=1}^n E\xi_{nk}^2 = 1$. Moreover, it follows that

$$\begin{aligned}
\sum_{k=1}^n E\xi_{nk}^2 I(|\xi_{nk}| > \delta) &= \sum_{k=1}^n {a'_{nk}}^2 EV_k^2 I(|a_{nk}V_k| > \delta) \\
&\le \sum_{k=1}^n \frac{a_{nk}^2}{\sigma_k^2} EV_k^2 I(\max_{1\le k\le n} |a_{nk}V_k| > \delta) \\
&\le (\inf_k \sigma_k^2)^{-1} \sup_k E\{V_k^2 I(\max_{1\le k\le n} |a_{nk}V_k| > \delta)\}.
\end{aligned}$$

It follows from the condition (24) that

$$\sup_k E\{V_k^2 I(\max_{1\le k\le n} |a_{nk}V_k| > \delta)\} \to 0 \quad \text{as } n \to \infty.$$

Lemma A.4 is therefore derived from Lemma A.3.

REFERENCES

Blanchflower and Oswald (1994). Blanchflower, D.G. and Oswald, A.J. (1994). *The Wage Curve*, MIT Press Cambridge, MA.

Carroll, R.J. (1982). Adapting for heteroscedasticity in linear models, *Annals of Statistics*, **10**, 1224-1233.

Chen, H.(1988). Convergence rates for parametric components in a partly linear model. *Annals of Statistics*, **16**, 136-146.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory.* 2nd Edition, Springer-Verlag.

Cuzick, J (1992a). Semiparametric additive regression. *Journal of the Royal Statistical Society, Series B*, **54**, 831-843.

Cuzick, J (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Annals of Statistics*, **20**, 1129-1136.

Engle, R. F., Granger, C.W.J., Rice,J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.

Eubank, R.L., Hart, J.D. and Speckman, P.(1990). Trigonometric series regression estimators with an application to partially linear model. *Journal of Multivariate Analysis*, **32**, 70-83.

Gao, J. T., Hong, S.Y. and Liang, H. (1995). Convergence rates of a class of estimates in partly linear models. *Acta Mathematicae Sinica Series New*, **38**, 658-669.

Hall, P. and Carroll, R.J. (1989) Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society, Series B*, **51**, 3-14.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, New York.

Heckman, N.E.(1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.

Liang, H. (1995). A note on asymptotic normality for nonparametric multiple regression: the fixed design case. *Soo-Chow Journal of Mathematics,* 395-399.

Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics*, **15,** 610-625.

Rendtel, U. and Schwarze, J. (1995). *Zum Zusammenhang zwischen Lohnhoehe und Arbeitslosigkeit: Neue Befunde auf Basis semiparametrischer Schaetzungen und eines verallgemeinerten Varianz-Komponenten Modells.* German Institute for Economic Research (DIW) Discussion Paper 118, Berlin, 1995.

Robinson, P.W. (1987). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica*, **55**, 875-891.

Schick, A. (1993). On efficient estimation in regression models. *Annals of Statistics*, **21**, 1486-1521.

Schick, A. (1996). Weighted least squares estimates in partly linear regression models. *Statistics & Probability Letters*, **27**, 281-287.

Speckman, P.(1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.

Willis, R. J. (1986). *Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions in: Ashenfelter, O. and Layard, R. The Handbook of Labor Economics*, Vol.1 North Holland-Elsevier Science Publishers Amsterdam, 1986, pp 525-602.